



TITLE:

京都大学におけるデータマネジ メントの現状: アンケート調査・ヒア リング調査報告

AUTHOR(S):

川口, 朋子; 家森, 俊彦; 元木, 環; 青木, 学聡; 梶田, 将
司

CITATION:

川口, 朋子 ...[et al]. 京都大学におけるデータマネジメントの現状: アン
ケート調査・ヒアリング調査報告. 2019

ISSUE DATE:

2019-02-28

URL:

<http://hdl.handle.net/2433/243957>

RIGHT:

主催者の許可を得て登録しています

2019.2.28 第2回京都大学リサーチデータマネジメント ワークショップ
理学研究科セミナーハウス

京都大学におけるデータマネジメントの現状

アンケート調査・ヒアリング調査報告

アカデミックデータ・イノベーションユニット
川口朋子 家森俊彦 元木環 青木学聡 梶田将司

アンケート調査 概要

- 次世代研究環境創成のための研究データに関する基礎調査
- 対象 京都大学に所属する研究者
教授・准教授・講師・助教・助手・特定有期雇用教員 3,375名（2018年5月1日現在）
- 方法
- 調査期間と回答方法
- 2018年12月4日-
先行調査 ユニットメンバーへメールで質問紙を送付－返信で回答
- 2018年12月21日-2019年1月15日
本調査：環境情報機構アンケートシステムに登録
- 調査告知方法
 - ・ 総務課経由のメールによる各部局への依頼
 - ・ 学際融合教育研究推進センターのメーリングリストへの配信
 - ・ 京大教職員用ポータルサイトの掲示板への掲載

次世代研究環境創成のための研究データに関する基礎調査 A Basic Survey of Research Data for Next-generation Research Support Environment

学内におけるデータの所在情報などを収集するためのアンケート調査を実施します。調査の対象となる研究データは、公式に提供、共有しているものに限らず、個人で収集、使用されているものも想定しております。最大5分程度の簡単な予備調査ですので、ぜひご自身の研究データについて、ご回答、ご協力をお願いいたします。

Please join this survey about research data in Kyoto University. We are supposing that the targeted research data is not only the case where it is published and shared officially, but also the one where it is collected and used in privately closed manner. This preliminary survey takes only about five minutes so that you can easily join using your spare time.

このアンケートは全部で 14 問あります。
回答者の情報

•
回答者氏名
Full name of respondent

• 回答者の所属情報(部局, 講座等)
Affiliation of respondent (department, labo name etc.)

• 回答者の職名
Position title of respondent

• 連絡先 e-mail アドレス
Contact e-mail address

• 研究データ(データセット、データベースもしくはデータベース化したいデータ)をお持ちですか。
Do you own a research data (dataset, database and/or a raw data to make database?)

☐ はい ☐ いいえ

研究データの詳細

先ほどの回答で「はい」とお答えになった場合、以下の内容をご記入ください。

質問 a.
データ、データセットもしくはデータリポジトリの名称はありますか? あればその名称をお答えください。
Question a.
Does your data (or dataset, data repository) have a name? Please write the name if it has.

質問 b.
データ取得・収集時の研究分野や想定した目的はどのようなものですか?
Question b.
What is the targeted research area and its purpose of collected or obtained data?

アンケート回答画面

データ取得・収集時の研究分野や想定した目的はどのようなものですか？
Question b.
What is the targeted research area and its purpose of collected or obtained data?

質問 c.
データ内容の概略。データの形式や数量などをお書きください。
Question c.
Abstract of data content. Please specify the format and amount of data.

質問 d.
データを表すキーワード(5件まで)
Question d.
Keywords of data (five at most)

キーワード1	<input type="text"/>
キーワード2	<input type="text"/>
キーワード3	<input type="text"/>
キーワード4	<input type="text"/>
キーワード5	<input type="text"/>

質問 e.
回答者とデータの責任者が異なる場合は、責任者の氏名／職名を教えてください。
Question e.
Please tell us the responsible person's name and title if the respondent and the responsible person is different.

質問 f.
データ公開の状況について教えてください。
Question f.
Please write the status of data publication.

以下から一つをお選び下さい。

- ☐ すでに公開している(下記の「データへのアクセス方法」に記入してください)
Already available to the public (if checked, put additional answer at question f "how to access it" below.)
- ☐ 公開を準備中である
Preparing for publication.
- ☐ 研究グループ関係者などある範囲内のみ共有、提供している
The data is shared among limited persons and groups, such as laboratory members or research community.
- ☐ 現在は公開していないが、研究終了後に公開を予定している
Currently, data is not open for public, but will be open when the research is finished.
- ☐ 公開予定はなく、研究終了後データは破棄する予定である
No plan for publication. The data should be disposed after the research.

質問 f.

データ公開の状況についてお教えてください。

Question f.

Please write the status of data publication.

以下から一つをお選び下さい。

- ☐ すでに公開している(下記の「データへのアクセス方法」に記入してください)
Already available to the public (if checked, put additional answer at question f' "how to access it" below.)
- ☐ 公開を準備中である
Preparing for publication.
- ☐ 研究グループ関係者などある範囲内のみ共有、提供している
The data is shared among limited persons and groups, such as laboratory members or research community.
- ☐ 現在は公開していないが、研究終了後に公開を予定している
Currently, data is not open for public, but will be open when the research is finished.
- ☐ 公開予定はなく、研究終了後データは破棄する予定である
No plan to publication. The data should be discarded after the research.
- ☐ 現在公開していない。今後も公開する予定はない。
The data is not open to the public. There is no plan to release it in the future.
- ☐ 公開したいが、課題を抱えているため実施できていない。
I want to release it, but I can not do it because I have a problem
- ☐ その他 :
- ☒ わからない

上記の回答内容を、調査結果として公開して良いものを選択してください。特記事項がある場合、「その他」にコメントしてください。

Please select your answers that we can publish as survey results. Please write comments in "Other" if you have specific notes.

あてはまるものをすべてチェックしてください。

- ☐ 質問a. データの名称
Question a. Title of the research data
- ☐ 質問b. データ取得の目的、学術分野
Question b. purpose and research area of your data
- ☐ 質問c. データの形式、分量
Question c. Content and quantity of the research data
- ☐ 質問d. データを表すキーワード
Question d. Keywords related with the research data
- ☐ 質問e. データの管理責任者情報
Question e. Contact to the responsible person of the research data
- ☐ 質問f. データの公開状況
Question f. Publication status of the research data
- ☐ 質問f'. 公開データへのアクセス
Question f'. Access to the published data
- ☐ その他 :

追記事項・コメント、データの管理、保管、提供などについて、お困りのことがございましたら自由にご記述ください。

Please feel free to write any comments if you are facing any difficulties about data management, data archive and data publication.

保存した入力内容の読み込み

後で再開

送信する

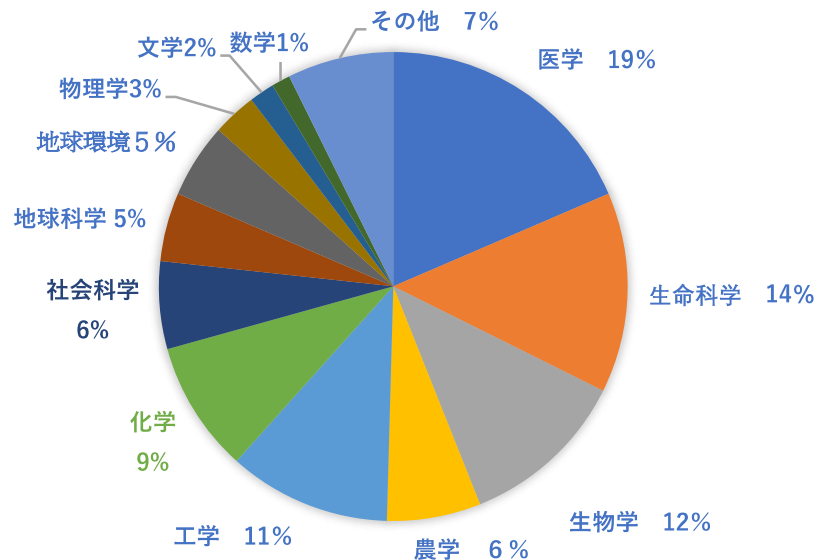
入力内容を消去して終了

アンケート 回答数と回答者の研究分野

- 全問回答数244 未完了回答584 合計828 (2018.02.23現在)



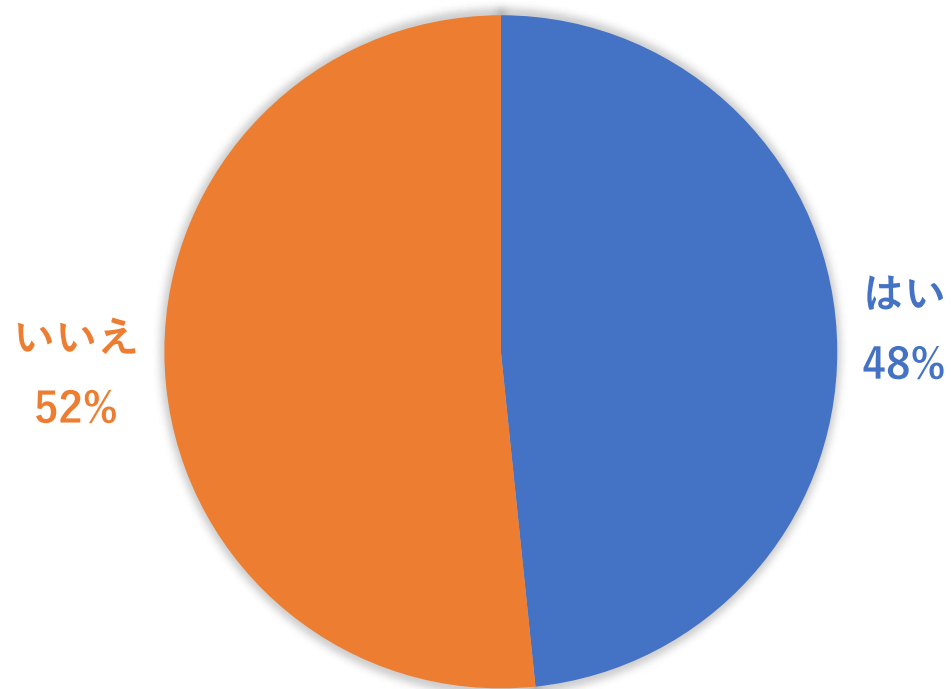
- 回答者所属機関の研究分野



分野	合計
医学(心理学を含む)	43
生命科学(ミクロ)	32
生物学(マクロ)	27
農学(マクロ)	15
工学(防災を含む)	26
化学	21
社会科学	14
地球科学(理学的)	11
地球環境(工学的)	12
物理学	7
文学	4
数学	3
学際的・その他	17

Q 1 研究データ(データセット、データベースもしくはデータベース化したいデータ)をお持ちですか。

はい	118
いいえ	126



Q2.

データ、データセットもしくはリポジトリの名称はありますか？あればその名称をお答えください

- ud-kanbun
- 中性子捕獲断面積
- Digital Morphology Museum, KUPRI
- 遺伝子多型を用いた 前立腺癌診断マーカーの開発
- ゲノム、トランスクリプトームデータ
- ゲノムネット<https://www.genome.jp/>、他様々な有名・無名のデータをもつ
- 合成化学, 量子化学計算, X線吸収分光
- UMIN ICDR
- 気象庁予報データ、気象庁長期再解析データ
- Mendeley Data
- 木材科学文献データベース
- 公式名称なし

Q3

データ取得・収集時の研究分野や想定した目的はどのようなものですか

- 研究で得たDNA情報を登録し、Accession Numberを受けること。研究で採集した生物標本とは別に、採集地情報などをエクセルファイルで京大博物館に分譲すること
- 洪水氾濫計算・浸水被害算定
- 森林科学
- 野生動物の個体数や密度などに関わる基礎情報
- 核データ、私のデータと以前のデータを比較するため
- 創業期の企業家の省察を明らかにすること
- 分野は心理学と神経科学で、取得・収録されたデータは、通常の解析およびデータマイニングの目的に使用される
- 研究分野：細胞生物学、目的：論文作成
- 環境保健、疫学研究の解析に用いる
- 収集分野：土壌学・環境農学
- 化合物半導体薄膜の作製条件
- 発達性ディスレクシアの神経基盤の解明 など

Q 4

データ内容の概略。データの形式や数量などをお書きください。

回答	119
わからない	130

- ・ fastaファイル、エクセルファイル
- ・ 地形データ（ASCII形式）、資産データ（テキスト形式、シェープファイル形式）、水文データ（テキスト形式）、気象データ（テキスト、バイナリ形式）、堤防断面データ（CAD形式）
- ・ 野帳や電子データの様々な形態、エクセルなどの表形式、GISのベクトルデータ、自動撮影カメラの写真（jpeg） 1 T B 程度
- ・ 実務家と研究者との対話の逐語記録
- ・ テキスト，動画，および数値解析ソフトウェア MATLAB で使用。mat 形式.
- ・ スプレッドシートデータ，写真，原稿，パワーポイントスライド，動画。20GB程度
- ・ tif画像、excelファイルなど（数量：不明）
- ・ エクセル・ファイルにまとめた世界各地の土壌分析データ
- ・ 世界の土壌断面や景観の写真、位置情報をまとめたパワポ・ファイル、数百点分
- ・ 質問紙調査
- ・ エクセルデータ 更新して古いのを含めると100以上 など

Q 5

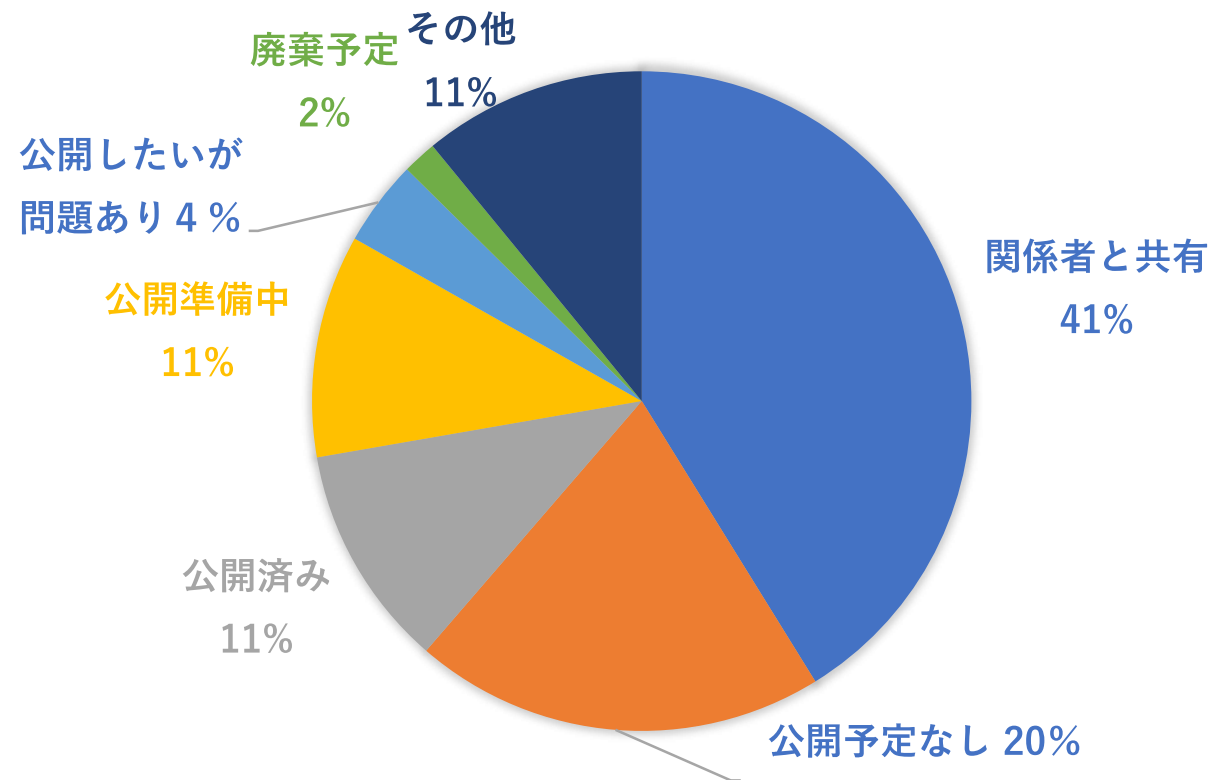
データを表すキーワード（5件まで）

- 洪水
- 野生動物
- サイズ
- 中性子エネルギー
- Landscape
- RNA-Seq
- 逐語記録
- 行動
- スプレッドシートデータ
- Health
- 炸裂条件
- 夢
- 発達性ディスクレシア
- Brain tumor
- 心臓再生医療
- 心理学実験
- 古典中国語
- 霊長類
- 前立腺癌
- フランス語教育
- eye imaginal discs
- パフォーマンス評価
- フェノロジー
- 微生物群集構造
- 骨格筋
- 気象
- 地盤災害
- X線解析
- molecular structure
- DNA
- 津波
- 作血幹細胞移植
- SAR
- 電子顕微鏡
- rheumatoid arthritis
- 抗うつ薬
- 多次元
- MRI など

Q 7

データ公開の状況について教えてください。

関係者と共有	49
公開予定なし	24
公開済み	13
公開準備中	13
公開したいが問題あり	5
廃棄予定	2
その他	13



自由記述

個人での保存・管理は負担が大きい

- 大量の画像データを管理するためのコストとセキュリティについて検討必要（農学）
- 画像はサイズが大きく保管場所に苦慮（化学）
- データ量が膨大（5TB/年）で積み重なり管理が大変。安心して保存できる場所が欲しい（生命科学）
- 10年保管して自動的に破棄するデータベースが欲しい（医学）
- 学内研究者だけがアクセスできるクラウドのような仕組みが京大内にあると便利だと思っています（理学）

体制的に継続的運用が困難

- サーバーの保守管理・更新に経費がかかり継続的な運用に不安。バックアップやセキュリティなど各種対応に現在の教員体制では難しい。京大内もしくは他の機関への移管も検討中（生物）
- 実験所は教員が一人しかおらず、継続的なデータ保存が問題。大学として継続的な保管方法・保管場所の確保が必要（理学）

自由記述

公開するには議論が必要

- 情報データ自体は提供元外部機関の所有物であり、個別契約を結んだ研究目的にのみ使用可能で、データそのものの公表や譲渡などは一切不可能。それらのデータから二次的に作成される研究成果物の所属は京都大学に属するが、個人情報などが関わるため一般公表はしていない。データ管理業務では、どこまでが外部団体等の所有物で、どこからが京都大学の所有物になるのかの区別を丁寧に考えていく必要があると思う。（農学）

公開作業を担う人間がない

- データベースキュレーターの存在が不可欠。この国に、この大学にこのような人を雇用する予算と意思があって初めてできること。（生命科学）
- 「データを保存すべき」の一声で、多様な学問領域や、日進月歩のデータ形式やメディアの現実を無視して、誰も守らないようなルールの策定は絶対にないようにして頂きたい。（生命科学）

自由記述

アンケートが回答しにくい（それぞれにとって「研究データ」の定義が異なる）

- 想定しているデータの意味が広すぎ。回答は困難（工学）
- 一つの質問の中に回答すべき内容が複数あったり質問が抽象的すぎて回答しにくかった（社会科学）
- 通常理系の実験室では多種多様なデータを持っている。質問の形式が回答しにくい（化学）
- 質問は自然科学を想定しており、人文科学にはあてはまらないと思う（文学）

アンケート調査 小括

1. データに関する認識が多様で曖昧
2. 学内に多種多様なデータが存在
3. 個人でのデータ管理は負担が大きい
4. 組織的、継続的なデータ管理が難しいケースを確認
5. 公開するには今後の議論が必要
6. 今後、実施方法に工夫が必要 → 確実に研究者にアンケートが届くには・・・

ヒアリング調査の概要

- 目的 アンケート調査の結果を深掘りする
- 期間 2018年12月26日~2019年2月20日
- 実施数 7名
- 方法 アンケート調査に回答された方
分野が偏らないよう配慮して抽出、依頼、スケジューリング
研究室を訪問し1時間程度のインタビューを実施

ヒアリング内容

1. 収集について

どんなデータか、なぜ集めているのか、いつから、誰が、どこで、
どうやってそのデータを集めているのか、課題など

2. 保管・整理について

どんな形態で整理・保管・分類しているのか（データベースの名称、
形態、作り方など）、保管後の利用目的、いつから、誰が（責任者）
どこで、どうやってそのデータを保管しているのか、課題など

3. 公開について

公開の目的/非公開の目的、いつから、だれが、どんな方法でデータを
公開しているのか、公開前に想定していた利用者・利用目的、
公開後の実際の利用者・利用目的、課題など

学内でのデータマネジメントの現状と課題に関するヒアリング調査

実施日 年 月 日 () : ~ : 場所

調査者

1. 収集について

どんなデータか（データの種類）

なぜ集めているのか（研究の目的）

いつから、誰が、どこで、どうやってそのデータを集めているのか

課題・困っている点

2. 整理・保管について

どんな形態で整理・保管・分類しているのか（データベースの名称、形態、作り方など）

なぜ保管するのか（保管後の利用目的）

いつから（どの段階で整理・保管を始めたか）、誰が（責任者）、どこで、どうやってそのデータを保管しているのか

バックアップを取るタイミングはどれくらいか

課題・困っている点

3. 活用について

なぜ公開するのか／公開しないのか（公開の目的／非公開の理由）

いつから、誰が、どんな方法でデータを公開しているのか（公開のレベル）

公開前：誰がどのように利用するのを想定したか

公開後：誰がどのように利用しているのか

課題・困っている点

ヒアリング調査 1 (2018.12.26実施)

- ・ 図書館情報学、東南アジア研究

収集	フィールドで実施したインタビューデータ(テキスト、音声)
整理	音声データはdropboxへ 現地の人に書き起こしを依頼 (インドネシア語、英語) →プリント、ファイルに綴じる
保管	ファイル
活用	論文執筆、インタビューーへのフィードバック

- ・ 取り扱いには常に不安がある。インタビューーへの配慮が重要
- ・ 公開は一部の研究者には利益がある。期間限定的に限定したグループ内への公開は意味があるかも。
- ・ インタビューデータについて「その人の利益にならないことは公開しない」という考えに賛同

ヒアリング調査 2 (2019.1.16実施)

- 大学法人文書など

収集	法人文書、各部局広報用刊行物、寄贈資料等の現物資料、文書のPDFや画像
整理	目録作成（識別番号・件名・作成年月日を付与） 人手不足 →データベース登録 一部資料はマイクロ撮影 人事データはデジタル化せず
保管	4000m ² の書庫（温湿度管理） データは情報環境機構のホスティングサービス利用
活用	原則公開（使用目的を問わず、報告の義務も課さない） 所蔵資料検索データベースで検索可、一部紙焼きで閲覧 博物館での展示へ貸し出しなど

- 分散所蔵されている資料群あり
- データのバックアップ体制は情報環境機構のサービスを利用しているという認識
- 館内で不定期に共有フォルダを更新
- HDDに入れたデータを館内に置いている

ヒアリング調査 3 (2019.2.6実施)

- 目的 中国古典語の形態素解析・依存文法解析

収集	漢籍の文法解析（白文に符号を付与）
整理	git labに登録
保管	自前で立ち上げたサーバーに保管、利用。共同研究チームで管理
活用	公開（常に登録情報を更新）

- ライセンスはパブリックドメイン
- 業務でデータベースをいくつも作ってきたが、アクセスの少ないものはタイミングを見て閉じないといけないと思う
- データの公開後の取り回しについても、考える必要があるのでは

ヒアリング調査 4 (2019.2.13実施)

- 学術分野毎の研究風土や価値観に関する比較調査

収集	ウェブ上でアンケート調査を実施した回答データ、テキスト、集計データなど 1800人のデータ（公開前提）
整理	エクセルに登録 ワードプレスで作業
保管	
活用	FTPでダウンロード可能 公開（話題性を上げることが目的）

ヒアリング調査 5 (2019.2.13実施)

- 地球物理学

収集	北極スバールバル諸島設置の観測機で取得したオーロラの画像データ 4秒に1枚撮影（1枚 512×512pixel）
整理	データは共同研究者のP Cにも自動的に送信、共有 同一フォーマットでサーバーに蓄積 目録なし
保管	3重のバックアップ（不定期実施） 観測機にバックアップ用の大容量ポータブルハードディスク。保守点検の際に 持ち帰る
活用	論文執筆 研究チームで合意が取れたものから限定的に公開可能 公開する場合1日単位

ヒアリング調査 6

(2019.2.20実施)

・有機化学

収集	天然、合成化合物の各種分析データ
整理	エクセルなど
保管	個人のPCで管理 研究室の50年分の研究ノートを別室でストック
活用	supporting informationとして公開。学会の投稿規定に、立証に必要なデータが提出されているかチェック項目あり

- ・ 化学において新規化合物を新規として認めてもらうためにデータを公開する必要性は当然のこと
- ・ 論文に使わなかったデータを公開することはない
- ・ 基本的には一般性、普遍性がないため過去（特に1980年代以前）のデータは使えない

ヒアリング調査 7

(2019.2.20実施)

- 地域研究（工学系）

収集	水質分析データ
整理	結果をエクセルでまとめデータベース化
保管	研究室で教員が管理 学生は卒業前にデータをフォルダに入れる
活用	論文執筆

- データ構造が複雑で、体系的なデータにするのは困難
 - 採水する頻度、場所、分析項目、統計方法、解析方法が人によって異なる
- 解析データを公開するには注釈文が必要←大きな労力

ヒアリング調査 7

(2019.2.20実施)

- 地域研究（工学系）

収集	トイレアンケート調査の回答データ（ベトナム語、タイ語、カンボジア語）
整理	結果をエクセルでまとめデータベース化
保管	研究室で教員が管理
活用	論文執筆

ヒアリング調査 その他のコメント

- 研究を開始した時の目的と結果は変わるもの。研究途中でデータ整理のことなど考えていない。思った通りに最後まできれいにデータをとることは難しい。
- 公開できるデータできないデータの線引きは必要ではないか。
- データによっては注釈文を付与して公開する必要がある。
- 公開作業を担う人員は不可欠。
- 研究データに関する教育は必要に応じて教授が学生に行っており負担が大きい。データ管理をシステムの構築に頼るのでなく、学生へ教育できる若手研究者が欲しい。

ヒアリング調査 小括

1. データ管理の実態は研究分野・研究テーマによって実に様々（関係者、共同研究者、所属学会の規定、人員体制、予算なども影響）
2. 目的は研究であり、データ管理が目的ではない
3. オープンを前提とすると、公開データ/非公開データを線引きする議論にコストが相当かかる（すべてオープンでなくてもよい、段階的な管理などに配慮が必要）

アンケート・ヒアリング調査報告 まとめ

1. データに関する認識が多様で曖昧なことが改めて浮き彫りになった。
(ヒアリングで、話を理解することも大変である)
2. 学内に多種多様なデータが存在するが、研究分野やテーマによって整理・保存・活用
の方法はそれぞれ異なる。整理されていないデータも多いと推測できる。
3. 個人・研究室でのデータ管理は負担が大きい。有効な支援を求める声もあり、組織
的・継続的な運用については検討する意義がある。
4. 公開には丁寧な議論が必要。公開がふさわしくないデータもある。非公開を含め段階
があってもよい。